EDS ACTIVITY

AMAZON PRODUCT DATASET

NAME :- SHREYA DILIP BABDE

ROLL NO :- ET1-17

PRN :- 202401070115

BATCH :- ET1

GOOGLE COLLAB LINK

https://colab.research.google.com/drive/1QSmAVRBJVO6k6wvDyff-GW8j6Dl3BkD8?usp=sharing

PROBLEM STATEMENT

Problem 1: First 5 products match the dataset head

Problem 2: Correctly identifies 1 out-of-stock item (Echo Show 8)

Problem 3: Average rating calculation is correct (4.58)

Problem 4: Correctly identifies Echo Show 8 as heaviest (1.2kg)

Problem 5: Discount values are calculated correctly (price × discount%)

Problem 6: Properly filters high-rated products (≥4.5)

Problem 7: Echo Dot has most reviews (68,500)

Problem 8: Final prices after discount are accurate

Problem 9: Correct category counts (2 Smart Speakers, etc.)

Problem 10: Correct average for Smart Speakers ($37.49)

Problem 11: Echo Flex has highest discount (25%)

Problem 12: Correct in-stock weight sum (0.63kg)

Problem 13: Correct price range ($105)

Problem 14: Properly identifies budget products (<$50)

Problem 15: Correct average reviews by category

Problem 16: Price-Rating correlation (0.46) is plausible

Problem 17: Correctly identifies Fire TV Stick as lightest (0.05kg)

Problem 18: Correctly shows all products are discounted (100%)

Problem 19: Correct median price for in-stock items ($44.99)

Problem 20: Correct price categorization (Low/Medium/High)

Commands + Code + Text

```python
import pandas as pd
import numpy as np

data = {
    'product_id': ['B08N5KWB9H', 'B07PGL8ZSY', 'B08L5QZPY6', 'B08N5LWCND', 'B08N5KWB9Z'],
    'name': ['Echo Dot (4th Gen)', 'Fire TV Stick', 'Kindle Paperwhite', 'Echo Show 8', 'Echo Flex'],
    'category': ['Smart Speaker', 'Streaming Device', 'E-Reader', 'Smart Display', 'Smart Speaker'],
    'price': [49.99, 39.99, 129.99, 129.99, 24.99],
    'rating': [4.7, 4.6, 4.8, 4.5, 4.3],
    'reviews': [68500, 43200, 28700, 18600, 9200],
    'in_stock': [True, True, True, False, True],
    'weight_kg': [0.3, 0.05, 0.18, 1.2, 0.1],
    'discount_%': [15, 10, 5, 20, 25]
}

df = pd.DataFrame(data)
print("Amazon Product Dataset:")
print(df)
print("\n1. First 5 products:")
print(df.head())
print("\n2. Out-of-stock items:", len(df[df['in_stock'] == False]))
print("\n3. Average rating:", round(df['rating'].mean(), 2))
print("\n4. Heaviest product:", df.loc[df['weight_kg'].idxmax()]['name'])
df['discount_value'] = df['price'] * (df['discount_%']/100)
print("\n5. Discount values:")
print(df[['name', 'discount_value']])
print("\n6. High-rated products (≥4.5):")
print(df[df['rating'] >= 4.5][['name', 'rating']])
print("\n7. Most reviewed product:", df.loc[df['reviews'].idxmax()]['name'])
df['final_price'] = df['price'] - df['discount_value']
print("\n8. Final prices after discount:")
print(df[['name', 'price', 'final_price']])
```

1s completed at 10:02 PM

Commands + Code + Text

```
print("\n5. Discount values:")
print(df[['name', 'discount_value']])
print("\n6. High-rated products (≥4.5):")
print(df[df['rating'] >= 4.5][['name', 'rating']])
print("\n7. Most reviewed product:", df.loc[df['reviews'].idxmax()]['name'])
df['final_price'] = df['price'] - df['discount_value']
print("\n8. Final prices after discount:")
print(df[['name', 'price', 'final_price']])
print("\n9. Products per category:")
print(df['category'].value_counts())
print("\n10. Avg Smart Speaker price: $",
      round(df[df['category'] == 'Smart Speaker']['price'].mean(), 2))
print("\n11. Max discount product:", df.loc[df['discount_%'].idxmax()]['name'])
print("\n12. Total in-stock weight (kg):",
      round(df[df['in_stock']]['weight_kg'].sum(), 2))
print("\n13. Price range: $", round(df['price'].max() - df['price'].min(), 2))
print("\n14. Budget products (<$50):")
print(df[df['price'] < 50][['name', 'price']])
print("\n15. Avg reviews by category:")
print(df.groupby('category')['reviews'].mean())
print("\n16. Price-Rating correlation:",
      round(df['price'].corr(df['rating']), 2))
print("\n17. Lightest product:", df.loc[df['weight_kg'].idxmin()]['name'])
discounted = len(df[df['discount_%'] > 0])
print("\n18. Discounted products:", round((discounted/len(df))*100, 2), "%")
print("\n19. Median price (in-stock): $",
      round(df[df['in_stock']]['price'].median(), 2))
df['price_category'] = pd.cut(df['price'], bins=[0,50,100,200],
                              labels=['Low','Medium','High'])
print("\n20. Price categories:")
print(df[['name', 'price', 'price_category']])
```

1s completed at 10:02 PM

OUTPUT

Commands + Code + Text

```
Amazon Product Dataset:
   product_id                name          category   price  rating  reviews  \
0  B08N5KWB9H  Echo Dot (4th Gen)     Smart Speaker   49.99     4.7    68500
1  B07PGL8ZSY       Fire TV Stick  Streaming Device   39.99     4.6    43200
2  B08L5QZPY6   Kindle Paperwhite          E-Reader  129.99     4.8    28700
3  B08N5LWCND         Echo Show 8     Smart Display  129.99     4.5    18600
4  B08N5KWB9Z           Echo Flex     Smart Speaker   24.99     4.3     9200

   in_stock  weight_kg  discount_%
0      True       0.30          15
1      True       0.05          10
2      True       0.18           5
3     False       1.20          20
4      True       0.10          25

1. First 5 products:
   product_id                name          category   price  rating  reviews  \
0  B08N5KWB9H  Echo Dot (4th Gen)     Smart Speaker   49.99     4.7    68500
1  B07PGL8ZSY       Fire TV Stick  Streaming Device   39.99     4.6    43200
2  B08L5QZPY6   Kindle Paperwhite          E-Reader  129.99     4.8    28700
3  B08N5LWCND         Echo Show 8     Smart Display  129.99     4.5    18600
4  B08N5KWB9Z           Echo Flex     Smart Speaker   24.99     4.3     9200

   in_stock  weight_kg  discount_%
0      True       0.30          15
1      True       0.05          10
2      True       0.18           5
3     False       1.20          20
4      True       0.10          25

2. Out-of-stock items: 1

3. Average rating: 4.58

4. Heaviest product: Echo Show 8

5. Discount values:
                 name  discount_value
0  Echo Dot (4th Gen)          7.4985
1       Fire TV Stick          3.9990
2   Kindle Paperwhite          6.4995
3         Echo Show 8         25.9980
4           Echo Flex          6.2475

6. High rated products (>4.5):
```

✓ 1s completed at 10:02 PM

Commands + Code + Text

```
6. High-rated products (≥4.5):
                 name  rating
0  Echo Dot (4th Gen)     4.7
1       Fire TV Stick     4.6
2   Kindle Paperwhite     4.8
3         Echo Show 8     4.5

7. Most reviewed product: Echo Dot (4th Gen)

8. Final prices after discount:
                 name   price  final_price
0  Echo Dot (4th Gen)   49.99      42.4915
1       Fire TV Stick   39.99      35.9910
2   Kindle Paperwhite  129.99     123.4905
3         Echo Show 8  129.99     103.9920
4           Echo Flex   24.99      18.7425

9. Products per category:
category
Smart Speaker       2
Streaming Device    1
E-Reader            1
Smart Display       1
Name: count, dtype: int64

10. Avg Smart Speaker price: $ 37.49

11. Max discount product: Echo Flex

12. Total in-stock weight (kg): 0.63

13. Price range: $ 105.0

14. Budget products (<$50):
                 name  price
0  Echo Dot (4th Gen)  49.99
1       Fire TV Stick  39.99
4           Echo Flex  24.99

15. Avg reviews by category:
category
E-Reader            28700.0
Smart Display       18600.0
Smart Speaker       38850.0
Streaming Device    43200.0
Name: reviews, dtype: float64

16. Price-Rating correlation: 0.46
```

```
12. Total in-stock weight (kg): 0.63

13. Price range: $ 105.0

14. Budget products (<$50):
                 name  price
0  Echo Dot (4th Gen)  49.99
1       Fire TV Stick  39.99
4           Echo Flex  24.99

15. Avg reviews by category:
category
E-Reader            28700.0
Smart Display       18600.0
Smart Speaker       38850.0
Streaming Device    43200.0
Name: reviews, dtype: float64

16. Price-Rating correlation: 0.46

17. Lightest product: Fire TV Stick

18. Discounted products: 100.0 %

19. Median price (in-stock): $ 44.99

20. Price categories:
                 name   price price_category
0  Echo Dot (4th Gen)   49.99            Low
1       Fire TV Stick   39.99            Low
2   Kindle Paperwhite  129.99           High
3         Echo Show 8  129.99           High
4           Echo Flex   24.99            Low
```